

Projets MODAL MAP 474D  
Simulation Numérique Aléatoire (SNA)  
autour des évènements rares  
Promotion X2014

*Equipe pédagogique :*

F. Benaych-Georges, M. Bompaire, S. De Marco,  
G. Fort, E. Gobet, I. Kortchemski

*Période : Avril 2016 - Juin 2016*

*(Version préliminaire du 16 avril 2016 )*

## Table des matières

<b>1</b>	<b>Risques extrêmes en finance</b>	<b>4</b>
<b>2</b>	<b>Sûreté d'un réacteur nucléaire</b>	<b>7</b>
<b>3</b>	<b>Dynamique de population, processus de Galton-Watson</b>	<b>10</b>
<b>4</b>	<b>Sûreté aérienne</b>	<b>13</b>
<b>5</b>	<b>Ruines en assurance</b>	<b>16</b>
<b>6</b>	<b>Messages twitter</b>	<b>20</b>
<b>7</b>	<b>Microstructure des prix financiers</b>	<b>24</b>
<b>8</b>	<b>Vie et mort d'une population, processus de Yule</b>	<b>28</b>
	<b>Références bibliographiques</b>	<b>32</b>

**Objectifs.** Il s'agit de réaliser un projet de modélisation et simulation sur un des thèmes applicatifs décrits ci-dessous, dans le contexte des événements rares.

Des références bibliographiques sont données dans le descriptif de chaque projet, ces documents sont accessibles sur le moodle du cours ou via la BCX. Prendre connaissance de ces références est important et fait partie du travail d'expérimentation et d'investigation du MODAL. Par cette démarche on attend des étudiants qu'ils :

- se familiarisent avec le thème proposé,
- en comprennent les enjeux de sorte à se poser les *bonnes questions*,
- développent des outils stochastiques de simulation pour mieux appréhender les événements rares (quantile, distribution conditionnelle à l'évènement, statistique dans de tels événements, scénario rare typique...). Ils auront en particulier à adapter à leur contexte les méthodes abordés en cours ou en TP.

Cela implique de l'autonomie et des initiatives.

Il est demandé de tester différentes approches, de les *comparer quantitativement et qualitativement*, avec un regard critique.

**Encadrement :** une permanence de suivi de projets est assurée chaque vendredi après-midi en **Amphi Faurre**, de 13h30 à 15h30.

**Livrables :** un rapport écrit, avec les codes de simulation en annexe.

**Outils informatiques :** il est recommandé de choisir **Python** comme langage de développement.

L'utilisation de **Scilab** est aussi possible. En cas de besoin en calculs particulièrement intensifs, on pourra recourir à **Java** ou **C/C++**.

**Soutenance :** elle est prévue **vendredi 10 juin 2016**, d'une durée de 45' (30' de présentation, 15' de questions).

**Barème de notation du projet :** le projet est noté sur 20 pts.

- (4pts) Capacités d'initiatives/autonomie face aux difficultés/problèmes soulevés par le projet.
- (4pts) Capacités à décomposer le problème en plusieurs étapes intermédiaires, avant de parvenir à une version aboutie (démarche d'expérimentation).
- (5pts) Capacités à mesurer les améliorations, les apports, comparaison numérique.

- (4pts) Rapport écrit.
- (3pts) Soutenance orale.

Les 3 premiers items sont principalement notés pendant les séances de TPs.

# 1 Risques extrêmes en finance

Les produits dérivés sont des *produits financiers*, définis à partir d'un autre produit financier plus simple comme par exemple une action, un indice, une devise, une matière première ou un taux d'intérêt.

On trouve l'existence des produits dérivés à l'époque du pharaon Akhé-naton (XIV<sup>e</sup> s. av. JC), pour garantir à l'avance l'achat et la vente de blé. Ce n'est que dans les années 1970 que les marchés financiers se sont vraiment organisés pour échanger de tels produits, les banques endossant par exemple les risques de fluctuation des taux de change à la place d'entreprises internationales leur ayant acheté un produit d'assurance ad hoc.

Depuis une quinzaine d'années, sous l'impulsion du Comité de Bâle, se développe une vision globale des risques au niveau bancaire. Cette vision globale a été adoptée dès 1998, avec la Value At Risk (VaR), indicateur de risque mesurant le seuil de *pertes potentielles* à un horizon donné et pour un quantile donné (typiquement 99% à horizon 10 jours, ou 99.9% ...). Cet indicateur réglementaire est relié au montant de fonds propres à immobiliser en regard des risques endossés, pour faire face à des pertes a priori rares. Ce projet se concentre sur l'évaluation des pertes extrêmes sous l'angle fréquence, sévérité et scénario les engendrant, en considérant en particulier des seuils à 99.99% voire plus.

## • Portefeuilles d'options et modèles gaussiens

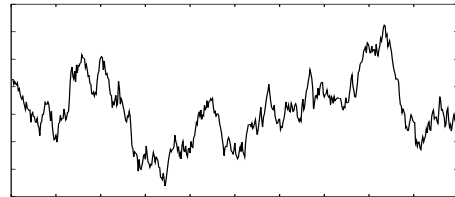
Le produit dérivé de base est le *Call* (option d'achat) donnant le droit (et non l'obligation) à son détenteur d'acheter à une date future  $T$  (*maturité*, de l'ordre de 6 mois à 1 an) une action de cours  $(S_t)_{t \geq 0}$  à un prix d'exercice  $K$  fixé à l'avance. A la date  $T$ , cela conduit à un équivalent financier égal à  $(S_T - K)_+$ . De manière analogue, le *Put* (option de vente) donne un équivalent financier égal à  $(K - S_T)_+$ .

On supposera que le cours d'action est modélisé par un mouvement brownien géométrique

$$S_t = S_0 e^{-\frac{1}{2}\sigma^2 t + \sigma W_t}.$$

Précisons les notations :

- $W = (W_t)_{t \geq 0}$  est un mouvement brownien, c'est-à-dire un processus continu à temps continu, partant de  $W_0 = 0$ , à accroissements indépendants et stationnaires, dont les accroissements  $W_{t+h} - W_t$  sont de loi gaussienne  $\mathcal{N}(0, h)$ .



- Le rendement de  $S$  (variation relative) sur un court intervalle de temps  $h \ll 1$  peut s'écrire  $\frac{S_{t+h} - S_t}{S_t} \approx \sigma(W_{t+h} - W_h)$ . Ainsi,  $\sigma$  mesure l'écart-type des rendements (appelé volatilité).

Dans ce modèle négligeant les taux d'intérêt, le prix  $\text{Call}(t, S_t)$  de l'option d'achat à la date  $t$  avant maturité  $T$  lorsque le sous-jacent vaut  $S_t$  est donné par la formule de Black & Scholes :

$$\text{Call}(t, S_t) = S_t \mathcal{N}[d_+(T - t, S_t, K)] - K \mathcal{N}[d_-(T - t, S_t, K)]$$

avec  $d_{\pm}(\tau, x, y) = \frac{1}{\sigma\sqrt{\tau}} \ln\left(\frac{x}{y}\right) \pm \frac{1}{2}\sigma\sqrt{\tau}$ , où  $\mathcal{N}$  est la fonction de répartition de la loi normale, centrée réduite. Le prix  $\text{Put}(t, S_t)$  du Put a une forme symétrique  $\text{Put}(t, S_t) = K \mathcal{N}[d_+(T - t, K, S_t)] - S_t \mathcal{N}[d_-(T - t, K, S_t)]$ .

On considère un portefeuille combinant des calls/puts sur  $I_0$  actions différentes. La valeur  $V_t$  du portefeuille est donc

$$V_t = \sum_{i=1}^{I_0} (\alpha_i \text{Call}_i(t, S_t^i) + \beta_i \text{Put}_i(t, S_t^i)).$$

On prendra par exemple

1.  $I_0 = 10$ ,  $\alpha_i = -10$ ,  $\beta_i = -5$ , pour tout  $i$  ;
2.  $I_0 = 10$ ,  $\alpha_i = 10$ ,  $\beta_i = 5$ , pour tout  $i$  ;
3.  $I_0 = 10$ ,  $\alpha_i = -10$ ,  $\beta_i = -5$  pour  $i = \{1, \dots, 5\}$  et  $\alpha_i = 10$ ,  $\beta_i = 5$  pour  $i = \{6, \dots, 10\}$ .

**Evolution indépendante.** Supposons d'abord que les mouvements browniens modélisant les actions sont indépendants.

▷ QUESTIONS À TRAITER :

1. évaluer numériquement les VaR et Conditional VaR (avec un seuil à 99.99% ou 99.999%),
2. évaluer la distribution conditionnellement à être au delà de la VaR,
3. influence de la composition du portefeuille.

**Evolution dépendante.** Supposons maintenant que les mouvements browniens sont corrélés positivement ( $\rho \geq 0$ ) du fait de la relation linéaire  $W_t^i = \sqrt{\rho} B_t^0 + \sqrt{1 - \rho} B_t^i$ , les  $(B^i)_{0 \leq i \leq I_0}$  étant des mouvements browniens indépendants.  $B^0$  peut s'interpréter comme un facteur commun à toutes les actions du marché.

▷ Reprendre les questions précédentes, tout en identifiant l'impact de la corrélation. On cherchera aussi à exhiber des trajectoires typiques de  $B^0$  associées aux pertes extrêmes.

## ● Extension à un modèle avec sauts

A la lumière de la crise, les pertes extrêmes apparaissent lors de chocs dans le marché, chocs qui se traduisent par des sauts négatifs sur les cours boursiers (d'amplitude de -5% à -10% par exemple). Il est alors assez naturel de superposer au mouvement brownien  $B^0$  un processus de Poisson composé avec des sauts négatifs, pour modéliser ces chocs. On abordera les questions précédentes dans ce nouveau modèle.

## ● Références bibliographiques

Sur les options et les modèles stochastiques en finance : [LL91], [TT3A].

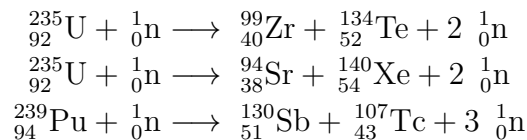
Sur les mesures de risque en finance : [FS02].

Sur le calcul de risques extrêmes par simulations : [GHS00], [GHS02], [Gla03], [CC10].

## 2 Sûreté d'un réacteur nucléaire

Le milieu du XX<sup>e</sup> siècle a marqué le début du développement du nucléaire aux Etats-Unis, en URSS, en France . . . voir <http://nucleaire.cea.fr/fr/repere/historique.htm>. En dehors des applications militaires (bombe atomique, propulsion nucléaire navale), les applications civiles concernent surtout la production d'électricité via les centrales nucléaires. Depuis les années 1960, les technologies ont considérablement évolué, utilisant surtout le principe de fission alors que les projets à l'étude concernent la fusion (ITER par exemple).

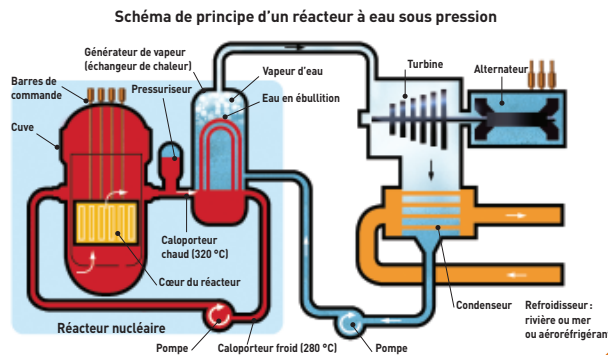
La fission nucléaire, découverte à la fin des années 30, est le phénomène de fragmentation de noyaux lourds (Uranium 235 ou 238, Plutonium 239) par capture de neutrons, donnant lieu à des nucléides plus légers et produisant une énergie importante. Par exemple



La fission produit de nouveaux neutrons qui peuvent entraîner une réaction en chaîne. L'énergie thermique dégagée est ensuite transformée (production de vapeur) en énergie mécanique ou électrique. Pour permettre à la réaction de continuer, il faut une quantité suffisante de combustible nucléaire dans le réacteur. A l'inverse, pour éviter un emballement de la réaction, des barres de commande peuvent être introduites dans le cœur du réacteur : ces barres constituées de matériaux spéciaux ont pour effet d'absorber les neutrons et ainsi d'arrêter la réaction.

L'accident de Three Mile Island en 1979 et la catastrophe de Tchernobyl en 1986 avaient déjà rappelé les risques du nucléaire. La catastrophe de Fukushima en 2011 a relancé le débat sur l'évaluation des risques extrêmes, dans un contexte où nos sociétés s'interrogent sur nos modèles de dépendance énergétique<sup>1</sup> et le développement d'énergies renouvelables.

Dans une centrale nucléaire, le danger le plus important provient d'un défaut de refroidissement du cœur du réacteur (avec pour conséquence un risque



Source : <http://www.cea.fr/content/download/4687/27890/file/livret-reacteur.pdf>

1. En France, la filière nucléaire fournit à elle seule 78% de la production électrique.

de fusion libérant les éléments très radioactifs). La prise en compte de tous les facteurs est complexe<sup>2</sup> et nous étudions un modèle simplifié, cherchant à évaluer la capacité des barres de commande à absorber les neutrons (arrêt d'urgence de la réaction), dans un contexte où les réflecteurs entourant le cœur sont endommagés (fissure, dégradation des propriétés réfléchissantes).

## ● Modèle neutronique

Nous modélisons le déplacement d'un seul neutron. Pendant son parcours (modélisé dans cette partie dans le plan), il est soumis à de nombreux chocs qui se produisent de manière imprévisible : à chaque choc, sa vitesse saute d'une quantité tirée aléatoirement et indépendamment de son parcours précédent. Si  $(\tau_i)_i$  est la suite des instants de chocs, sa vitesse (en km/s) est

$$V_t = V_{\tau_i^-} + Y_i, \quad \text{pour } t \in [\tau_i, \tau_{i+1})$$

où  $(Y_i)_i$  est une suite de v.a.i.i.d. dans  $\mathbb{R}^2$  modélisant les sauts de vitesse (chocs). La vitesse à la date  $t$  est donc :

$$V_t = V_0 + \sum_{i=1}^{N_t} Y_i$$

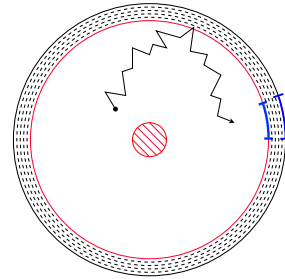
où  $N$  est un processus de Poisson d'intensité  $\lambda$  grande. Les variables  $(Y_i)_i$  et le processus  $N$  sont indépendants. La loi de  $Y$  est centrée et on pourra choisir une loi gaussienne sur chaque coordonnée, d'écart type  $2000 \text{ km s}^{-1}$  (neutron lent). Les ordres de grandeur réalistes pour  $\lambda$  sont de l'ordre  $10^{12} \text{ s}^{-1}$  à  $10^{15} \text{ s}^{-1}$ .

Si l'on ne tient pas compte des barres de commande et de l'enceinte du réacteur, la position du neutron au cours du temps évolue donc comme

$$X_t = X_0 + \int_0^t V_s ds.$$

C'est un modèle simple d'*équation de transport*, voir [LPS98]. Le cœur du réacteur est composé d'une enceinte extérieure réfléchissante (qu'on suppose circulaire de diamètre  $D = 5m$ ), et au centre d'une barre de commande (de diamètre  $d$ ). Plusieurs situations peuvent se présenter.

- Lorsque le neutron touche la barre de commande, il est absorbé et s'arrête.



Modèle de réacteur et déplacement de neutron

2. voir la conférence [http://www.coriolis.polytechnique.fr/Confs/Hervieu\\_conf.pdf](http://www.coriolis.polytechnique.fr/Confs/Hervieu_conf.pdf)



- Lorsque le neutron atteint l'enceinte réfléchissante, il est réfléchi vers l'intérieur.
- Une partie de l'enceinte est endommagée, sur une portion angulaire mesurant quelques mm. On peut imaginer que si le neutron atteint cette partie, il quitte le cœur du réacteur.

On cherche à évaluer la probabilité de la perte de contrôle du neutron décrit par l'évènement

$A = \{\text{le neutron atteint la zone endommagée avant d'être absorbée par la barre}\}.$

▷ QUESTIONS À TRAITER :

- Approximation asymptotique du modèle lorsque l'intensité est très grande et application. On pourra chercher à remplacer le modèle avec un modèle gaussien.
- Evaluer  $\mathbb{P}(A)$  en fonction des positions et vitesses initiales.
- Déterminer la taille de la barre de commande pour maintenir une probabilité inférieure à  $10^{-5}$ .
- Déterminer la distribution du temps d'absorption conditionnellement à  $A^c$  ou du temps de sortie conditionnellement à  $A$ .

## ● Réacteur à plusieurs barres de commande

Le cœur du réacteur contient maintenant 4 barres de commande disposées en carré. A quelle distance du centre disposer les barres pour garantir une efficacité maximale ? Discuter de l'influence des autres paramètres.

## ● Références bibliographiques

Sur le nucléaire : <http://www.cea.fr/content/download/4687/27890/file/livret-reacteur.pdf>.

Site de l'Autorité de Sûreté Nucléaire : <http://www.asn.fr/index.php>.

Sur les équations de transport : [LPS98] ou [CDL<sup>+</sup>89].

Sur la neutronique : [BR78].

Sur les asymptotiques de diffusion : [Pap75].

### 3 Dynamique de population, processus de Galton-Watson

La dynamique des populations étudie la répartition et le développement quantitatif de populations d'individus. Elle s'intéresse aux mécanismes d'auto-régulation des populations, au problème de l'extinction ou de croissance d'une population ou à l'existence d'un éventuel état stationnaire ou quasi-stationnaire. Dans ce projet, pour simplifier nous considérons une seule population d'individus semblables, n'étant pas en compétition pour les ressources, sans interaction entre eux (pas de proie ni prédateur), sans processus de migration additionnel.

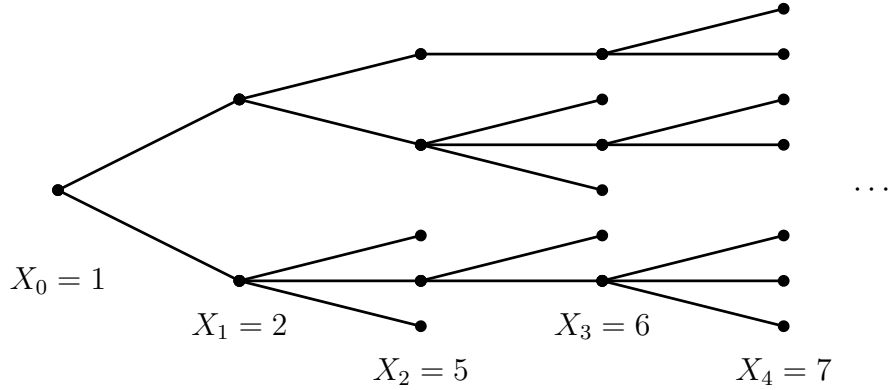


FIGURE 1 – Exemple d'arbre généalogique à partir d'un seul ancêtre

Un modèle simple et standard est le modèle de Bienaymé-Galton-Watson (BGW). On trouvera une analyse détaillée dans l'ouvrage [AN72]. Le modèle de BGW décrit l'évolution d'une population à travers l'évolution de la taille  $X_n$  de sa génération  $n$ . Avant de mourir, chaque individu  $k$  de la génération  $n$  donne naissance à  $Y_{n,k}$  individus, de sorte que

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{n,k}.$$

On suppose que les v.a.  $(Y_{n,k})_{n,k}$  sont indépendantes et de même loi  $\nu$ . Le processus  $(X_n)_{n \geq 0}$  est ainsi une chaîne de Markov à valeurs sur les entiers positifs. Pour que le problème ne soit pas dégénéré et un minimum intéressant, on supposera que  $0 < \mathbb{P}(Y = 0) < 1$  et  $\mathbb{P}(Y \geq 2) > 0$ . La taille initiale  $X_0$  de la population n'est pas forcément égale à 1.

Le point 0 est absorbant : si  $X_n = 0$  pour un  $n$ , alors  $X_p = 0$  pour tout  $p \geq n$ , autrement dit la population s'est éteinte. Le temps d'extinction est noté

$$T_E = \inf\{n \geq 0 : X_n = 0\}.$$

## • Loi de reproduction géométrique

On suppose ici que  $\nu = \mathcal{G}(m)$ , c'est-à-dire

$$\mathbb{P}(Y = k) = \frac{1}{m+1} \left(\frac{m}{m+1}\right)^k, \quad k \geq 0.$$

Le paramètre  $m > 0$  ci-dessus s'interprète comme la moyenne du nombre d'enfant d'un individu donné.

**Le cas sous-critique :**  $m < 1$ . Dans ce cas, *p.s.* la population s'éteint en temps fini. L'évènement de non-extinction  $A_n = \{X_n > 0\}$  à la génération  $n$  devient rare si  $n$  est grand.

▷ QUESTIONS À TRAITER :

1. estimer numériquement  $\mathbb{P}(A_n)$ ;
2. estimer la loi de  $X_n$  conditionnellement à la survie  $A_n$ ;
3. générer des scénarii typiques où la population n'est pas éteinte à la génération  $n$ .

L'influence jointe du nombre  $n$  de génération et de la taille initiale  $X_0$  de la population sera aussi étudiée. Les impacts de  $m$  et de la distribution  $\nu$  pourront être analysés. On pourra considérer également le problème des ressources consommées par la population, en évaluant  $\mathbb{P}(\max_{1 \leq i \leq n} X_i > x)$  ou  $\mathbb{P}(\sum_{1 \leq i \leq n} X_i > x)$  pour  $x$  grand.

**Le cas sur-critique :**  $m > 1$ . Dans ce cas, avec probabilité positive la population croît exponentiellement comme  $m^n$ . On s'intéressera au cas où la croissance est moins rapide, disons à la vitesse  $c^n$  où  $c \in ]1, m[$  : c'est un évènement rare lorsque  $n$  est grand.

▷ QUESTIONS À TRAITER :

1. estimer numériquement  $\mathbb{P}(0 < X_n \leq c^n)$ ;
2. quel scénario conduit à de telles croissances ?

## • Extension à une loi de reproduction aléatoire

Maintenant on prend en compte le fait que la reproduction de la population dépend de son environnement, qui change aléatoirement à chaque génération. A la génération  $n$ , la loi du nombre d'enfants est encore de type géométrique avec un paramètre  $m_n$ , qui est lui-même une variable aléatoire prenant deux valeurs  $m_\heartsuit$  (environnement favorable) et  $m_\dagger$  (défavorable). Les v.a.  $(m_n)_n$  sont supposées être i.i.d..

Dans ce contexte, a priori les comportements en temps long de la population peuvent être différents du cas précédent et il s'agit de mieux comprendre quel rôle a l'environnement dans les fluctuations anormales. On pourra traiter les questions précédentes dans ce nouveau contexte.

## ● Références bibliographiques

Sur les bases des modèles d'évolution de population : [AN72], [M3A].

Sur les applications : [JY98], [LES03], [KAT<sup>+</sup>08].

Sur des aspects plus avancés de modélisation : [Ner77], [GKV03], [BB09].

## 4 Sûreté aérienne

D'après le code de navigabilité international réglementant les avions civils à turbine, la probabilité d'une défaillance conduisant à une catastrophe doit être inférieure à  $10^{-9}$  par heure de vol. Les défaillances peuvent être d'origine multiples et nous ne les analysons pas de manière exhaustive dans ce projet.

Nous nous concentrons sur la probabilité de collision entre deux avions, que nous souhaitons être aussi très faible pour atteindre un objectif de sécurité raisonnable en adéquation avec la réglementation. Le cas contraire pourra être le point de départ pour renforcer la fiabilité des instruments de pilotage, modifier les plans de vols... Améliorer la sûreté aérienne et minimiser le risque de collision peut être réalisé à plusieurs niveaux.

- A un niveau global, il s'agit de répartir au mieux les vols entre les différents aéroports et entre les différents moments de la journée/semaine.
- Au niveau le plus fin, lorsque la collision est imminente, cela relève du sang-froid des pilotes et de la réactivité des tours de contrôles.
- A un niveau intermédiaire et pendant le vol, il s'agit d'orienter les avions sur des routes qui restent assez distantes les uns des autres. Néanmoins, les trajectoires des avions ne sont pas déterministes car elles sont soumises à des aléas météorologiques (vent), à des perturbations de pilotage, à des erreurs de mesure, qui hélas rendent la collision possible.

Pour évaluer ce dernier risque de collision et appréhender le type de scénario y conduisant, nous modélisons le déplacement aléatoire des avions.

### Unités d'usage.

- Distance : 1 mile nautique (noté  $1\ nmi$ ) = 1.852 km.
- Vitesse : 1 knot (noté  $1\ kt$ ) = 1 nmi/h = 1.852 km/h.  
Ainsi, une vitesse de 500 kt est égale à 926 km/h (c'est à dire une vitesse standard d'avion de ligne).  
On a aussi  $1\ nmi/min = 60\ kt = 111.12\ km/h$ .

Typiquement, le déplacement d'un avion s'effectue selon un plan de vol, qui est constitué de points de passage (*waypoint*) à intervalle régulier (20 min). Pour simplifier, on considère uniquement des trajectoires planes. Ainsi, entre deux points de passage, la trajectoire *idéale* (non perturbée) de l'avion est rectiligne, à vitesse constante  $v$  : si on note  $x_t = (x_{a,t}, x_{c,t})$  la position<sup>3</sup> de l'avion au cours du temps dans un repère local centré sur le précédent point de passage, avec l'axe des abscisses vers le prochain point de passage, on a  $x_{a,t} = vt$  et  $x_{c,t} = 0$ . A partir de données de vol [PW05], considérons un

---

3.  $x_{a,}$  désigne la coordonnée *along-track* et  $x_{c,}$  la coordonnée *cross-track*.

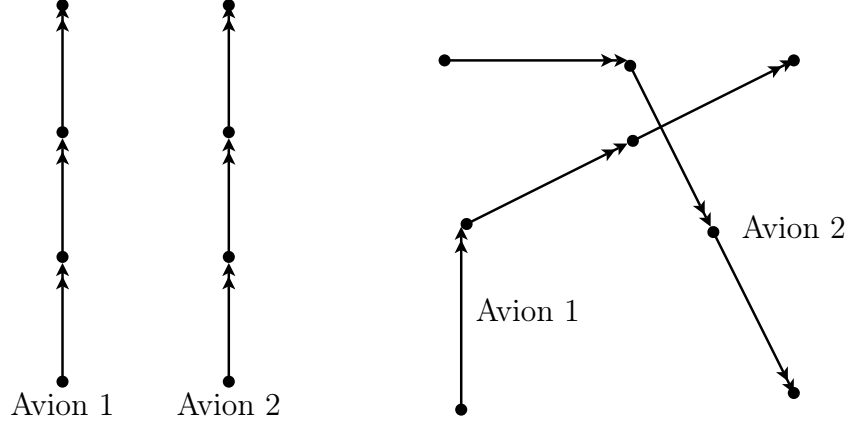


FIGURE 2 – Plan de vols parallèles et plan de vols se croisant. 3 points de passage (1 heure de vol).

modèle aléatoire  $(x_t)_{t \geq 0}$  en posant

$$x_{a,t} = vt + M_{a,t}, \quad x_{c,t} = M_{c,t},$$

où  $M_{a,.}$  et  $M_{c,.}$  sont deux processus gaussiens continus, indépendants, de fonction de variance/covariance donnée par (pour  $t < s$ )

$$\begin{aligned} K_a(t, s) &= \text{Cov}(M_{a,t}, M_{a,s}) = r_a^2 t^2, \\ K_c(t, s) &= \text{Cov}(M_{c,t}, M_{c,s}) = \sigma_c^2 (1 - e^{-2\frac{r_c}{\sigma_c} vt}) e^{-\frac{r_c}{\sigma_c} v(s-t)}. \end{aligned}$$

Les paramètres  $r_a = 0.25$  nmi/min,  $r_c = 1/57$  et  $\sigma_c = 1$  nmi sont communs à tous les avions civils à turbine. Ce qui peut changer d'un avion à un autre, c'est sa vitesse  $v$  et son plan de vol (la suite des points de passage).

Pour les besoins de simulations, les trajectoires ne seront pas considérées en temps continu, mais échantillonnées à intervalle régulier (par exemple 1 fois par minute) : c'est un paramètre de simulation à choisir convenablement. On s'attachera à détecter les *situations de conflit entre avions*, c'est-à-dire lorsque leur distance devient inférieure à 0.1 nmi, situation conduisant vraisemblablement à une collision imminente. La distance seuil pourra dépendre de la fréquence d'échantillonnage.

### Quelques explications sur le modèle.

- *Composante along-track.* Lorsque le temps augmente, la coordonnée *along-track* est de moins en moins certaine (cumul des aléas de navigation) : la croissance quadratique de la variance peut s'interpréter comme une incertitude sur la vitesse moyenne car  $x_{a,t} \in [(v - 1.96r_a)t, (v + 1.96r_a)t]$  avec probabilité 95%.

- *Composante across-track*. Sur la coordonnée *cross-track*, le phénomène est différent et la variance de l’incertitude se stabilise quand  $t$  est grand. Cela s’interprète ainsi : le long de la composante *along-track*, le pilote et les instruments de bord cherchent à corriger les écarts à la trajectoire idéale (ici 0) induisant un phénomène de retour à 0. Ce modèle porte le nom de processus d’Ornstein-Uhlenbeck. Ce processus a un comportement ergodique, ce qui se voit en calculant la corrélation pour des temps lointains ( $\lim_{s \rightarrow \infty} \text{Cor}(x_{c,t}, x_{c,s}) = 0$  avec vitesse exponentielle).

## ● Modèle à deux avions

On considère deux avions se déplaçant de manière indépendante, dans un même plan. Pour les tests, on considèrera différents plans de vol, comme par exemple celui de trajectoires parallèles, ou au contraire, celui des trajectoires qui se croisent (voir la figure 2). La distance entre deux points de passage correspond à un temps de vol prévu de 20 min. On supposera que chaque avion change de direction toutes les 20 minutes et se dirige vers son prochain point de passage.

▷ QUESTIONS À TRAITER :

1. Calcul numérique de la probabilité de conflit ; impact du plan de vol et des vitesses/distances entre avions ;
2. Distribution du temps de conflit, conditionnellement à l’occurrence du conflit ;
3. Distribution conditionnelle de la position de conflit.

On pourra dans un premier temps supposer aucune incertitude sur la composante *along-track* ( $r_a = 0$ ), puis considérer le cas  $r_a \neq 0$ .

## ● Application à l’optimisation de plan de vols

On considère un plan de vol du type du second de la figure 2 : les avions se croisent pour rejoindre chacun un coin d’un carré. Cette fois, on cherchera à déterminer les trajectoires les plus courtes possibles (par exemple pour minimiser la consommation carburant, ou le temps de vol) sous contrainte de probabilité de collision faible. L’optimisation à réaliser concerne uniquement les 4 waypoints intermédiaires.

## ● Références bibliographiques

Sur les processus gaussiens : [Nel67, Chap. 6 et 9].

Sur le calcul Monte Carlo des détections de conflit : [PW05].

Sur la modélisation des déplacements : [PE97], [GL04], [HPS05], [PW05].

Le site web du projet européen HYBRIDGE : <http://hosted.nlr.nl/public/hosted-sites/hybridge/>

## 5 Ruines en assurance

### ● Contexte

L'assurance est un contrat de transfert de risques entre deux parties : l'assureur prend à son compte un risque (sinistres, dommages à la personne...) en échange d'une cotisation versée par l'assuré. Les sociétés d'assurance s'appuient sur un grand nombre d'assurés pour mutualiser les risques (effet loi des grands nombres) ; en 2013<sup>4</sup>, le chiffre d'affaires de l'assurance en France dépasse les 180 milliards d'euros. L'évaluation de la solvabilité d'une société d'assurance est une question cruciale et complexe ; en 2008, le groupe américain AIG a dû être renfloué par la Réserve Fédérale Américaine alors qu'il était au bord de la ruine, son capital ayant été largement déprécié suite à la crise des *sub-primes*. En Europe, la réforme réglementaire Solvabilité II<sup>5</sup> impose de mieux mesurer les expositions aux risques, en évaluant par exemple les probabilités d'insolvabilité et de ruine. Afficher des excellents indicateurs de solvabilité sert aussi d'argument marketing auprès des clients.

Plus précisément, une compagnie d'assurance dispose d'un capital (fonds propres) ; elle perçoit régulièrement des cotisations de ses clients, et les indemnise en cas de sinistres. Cette compagnie propose en général des contrats de différents types liés aux différentes branches de son activité : assurance vie, complémentaire santé, assurance habitation, assurance automobile, assurance responsabilité civile, etc., avec des taux de cotisation qui dépendent de la *fréquence* des sinistres et de leur *sévérité*. Elle est ruinée lorsque les montants à rembourser sont supérieurs au capital de départ cumulé avec les cotisations perçues ; cela se produit en cas d'apparition de sinistres plus fréquents ou plus importants (catastrophes naturelles, épidémies...) ou si le capital initial et les cotisations sont insuffisants. Il est donc crucial de déterminer et/ou d'ajuster ces paramètres pour garantir la pérennité d'une compagnie d'assurance. Enfin, il est possible qu'une branche de l'assureur soit au bord de la ruine sans que l'assureur le soit globalement.



### ● Modélisation

Pour commencer, on considère une seule branche d'activité. La compagnie d'assurance a un capital de départ  $x$ , et on note  $R_t$  sa réserve à l'instant  $t$ . La compagnie d'assurance :

---

4. [http://www.ffsa.fr/sites/jcms/p1\\_983486/1-assurance-francaise-demonstre-sa-solidite?cc=fn\\_7345](http://www.ffsa.fr/sites/jcms/p1_983486/1-assurance-francaise-demonstre-sa-solidite?cc=fn_7345)

5. [https://fr.wikipedia.org/wiki/Solvabilité\\_II](https://fr.wikipedia.org/wiki/Solvabilité_II)



- perçoit des cotisations de ses clients, pour simplifier en continu et uniformément réparties sur l'année. On note  $p$  le taux de cotisation par unité de temps. Les recettes de la compagnie pendant un temps  $t$  sont donc  $pt$ .
- verse des indemnités à ses assurés sinistrés en fonction des dommages subis. On note  $0 = T_0 < T_1 < T_2 < \dots$  les instants aléatoires des sinistres et  $(Z_k)_{k \geq 1}$  les dommages. En posant  $N_t \stackrel{\text{def}}{=} \sum_{k \geq 1} \mathbf{1}_{T_k \leq t}$  le nombre de sinistres entre l'instant 0 et  $t$ , la réserve à la date  $t$  est

$$R_t = x + pt - \sum_{k=1}^{N_t} Z_k.$$

- ▷ **Modèle, paramètres.** On modélise les montants remboursés  $(Z_k)_{k \geq 1}$  par des lois Gamma i.i.d. de paramètres  $(\alpha, \beta)$ ,  $\alpha > 0$  et  $\beta > 0$ . Les durées entre deux sinistres successifs  $(\tau_k)_{k \geq 1}$  définies par  $\tau_k \stackrel{\text{def}}{=} T_k - T_{k-1}$ , sont supposées indépendantes entre elles et indépendantes des sinistres ; on suppose qu'elles sont distribuées suivant des variables exponentielles de paramètre  $\lambda$ ,  $\lambda > 0$ .
- ▷ **Ruine.** La société est ruinée à l'horizon  $T$  si la réserve  $R_T$  est négative.

## ● Risque de ruine à l'horizon 1 an

L'objectif de cette partie est de calculer la probabilité de ruine  $\mathbb{P}(R_T \leq 0)$  pour différentes valeurs des paramètres du modèle  $(x, \alpha, \beta, \lambda, p)$  et pour  $T = 1$  an.

1. Estimer la probabilité de ruine au bout d'un an pour les deux jeux de paramètres suivants  $\{p = 14, \alpha = 2.5, \lambda = 2, \beta = 0.5, x = 100\}$  et  $\{p = 14, \alpha = 3, \lambda = 1/24, \beta = 0.04, x = 250\}$ . Proposer différentes méthodes de calcul.
2. Déterminer le capital initial que la compagnie doit avoir pour que sa probabilité de ruine à l'horizon annuel soit inférieure à  $10^{-6}$ .
3. Estimer la distribution de  $R_T$  sachant qu'il y a eu ruine à l'instant  $T$ .

## ● Extensions

On traitera l'une des extensions suivantes.

- ▷ **Risque de ruine à horizon infini.** On veut quantifier si la compagnie d'assurance sera un jour ruinée. Si la société d'assurance est effectivement ruinée, la ruine interviendra à l'instant d'arrivée d'un sinistre : en conséquence, on s'intéresse au processus à temps discret de la somme déboursée échantillonnée aux instants d'arrivée des sinistres. La somme déboursée à l'instant

d'arrivée du  $n$ -ième sinistre est égale à  $X_1 + \dots + X_n$ , où  $X_k \stackrel{\text{def}}{=} Z_k - p\tau_k$ . La probabilité de ruine à horizon infini est donc donnée par

$$\mathbb{P}(\sigma_x < \infty), \quad \text{où} \quad \sigma_x \stackrel{\text{def}}{=} \inf\{n \geq 1, \sum_{k=1}^n X_k > x\},$$

avec la convention  $\inf \emptyset = +\infty$ .

La difficulté dans la simulation en temps infini provient du fait qu'il faut simuler toute la trajectoire  $(\sum_{k=1}^n X_k, n = 0, \dots, +\infty)$  pour déterminer si la réalisation est telle que  $\sigma_x < +\infty$  ou  $\sigma_x = +\infty$ . Pour pallier à cette difficulté, on propose d'effectuer un changement de probabilité exponentiel décrit dans [Gla03] et [BLG09] pour lequel sous une nouvelle probabilité  $\mathbb{P}_\theta$ ,  $\mathbb{P}_\theta(\sigma_x < \infty) = 1$ .

**Quelques résultats théoriques.** On note  $\psi_X$  la fonction génératrice des cumulants de  $X$ . Le livre [Asm03] et l'article [BLG09] rappellent des nombreuses propriétés de la fonction génératrice des cumulants. Si  $\mathbb{E}[X] < 0$ , il existe une unique valeur positive  $\theta^*$  telle que  $\psi_X(\theta^*) = 0$  : on a alors

$$\mathbb{P}(\sigma_x < +\infty) = \mathbb{E}_{\theta^*} \left[ e^{-\theta^* \sum_{k=1}^{\sigma_x} X_k} \right],$$

où  $\mathbb{E}_\theta$  désigne l'espérance associée à la probabilité  $\mathbb{P}_\theta$ . Voir [Gla03].

**Questions.** On cherchera à

1. estimer la probabilité de ruine pour  $\alpha = 1$ ,  $\beta = 1$ ,  $\lambda = 0.5$ ,  $p = 1$  et différentes valeurs de  $x$  avec la méthode présentée ci-dessus.
2. démontrer certains des résultats énoncés dans cette section.

▷ **Compagnie diversifiée.** Dans cette section, on suppose que la compagnie étudiée propose des assurances habitations et des assurances automobiles, et qu'une partie des sinistres liés aux risques naturels touchent à la fois le secteur automobile et habitation. On ajoute des sinistres propres aux habitations  $N_t^H$  et des sinistres propres aux voitures  $N_t^A$  : le nombre de sinistres affectant la branche habitation jusqu'au temps  $t$  est donc  $N_t^{RN} + N_t^H$ , et le nombre des sinistres affectant la branche automobile est  $N_t^{RN} + N_t^A$ .

1. Estimer le risque de ruine au bout d'un an (on pourra notamment étudier la ruine de chaque branche et les effets de compensation entre branche).
2. Déterminer le capital initial dont doit disposer la compagnie pour que sa probabilité de ruine à horizon 1 an soit inférieure à  $10^{-6}$ . Comment répartir un capital donné entre ses deux branches ?
3. Représenter la distribution de la réserve sachant qu'il y a eu ruine.

## ● Références bibliographiques

Théorie de la ruine : [Asm03].

Changement de probabilité exponentiel : [Gla03][BLG09].

## 6 Messages twitter



Twitter est une plateforme de microblogage, permettant d'envoyer rapidement de brefs messages (pas plus de 140 caractères), appelés **tweets** ("gazouillis"), sur internet. Cette plateforme est devenue très populaire, et a atteint les 500 millions d'utilisateurs en février 2012.

Ces messages contiennent (parmi d'autres) plusieurs informations : le moment de l'envoi du message (sous la forme d'un timestamp UNIX <sup>6</sup>), l'auteur et le corps du message. Certains messages forment des conversations, d'autres sont des messages décrivant un événement venant d'arriver dans le monde. Le nombre très élevé d'utilisateurs de twitter en fait une source d'information en temps réel très intéressante.

De nombreux tweets sont par exemple envoyés par des acteurs des marchés financiers. Quelques articles prétendent même que twitter permettrait de mieux prédire le comportement d'un marché financier que des analyses plus classiques <sup>7 8</sup>. Un exemple de tweet est le suivant :

```
User:  SchwartzNow
Time:  2009-12-17T19:20:15
Body:  also for tomorrow expect high volume options
       traded stocks like $aapl,$goog graviate around
       the strikes due to the delta hedging
```

Certains mots portent le sens du message : **tomorrow, high volume, options, stocks, \$aapl, \$goog**. Ce tweet contient des mots commençant par le symbole \$, appelés des "cashtags" : \$aapl et \$goog (AAPL et GOOG sont les tickers <sup>9</sup> des actions Apple et Google). Un exemple d'un tel jeu de données est disponible ici <sup>10</sup>.

### ● Objectifs

On veut construire un modèle simple pour le flux de tweets observés sur une période de temps. Les phénomènes observés dans ce type de flux, que l'on souhaite incorporer au modèle, sont les suivants :

---

6. Le timestamp UNIX correspond, par convention, au nombre de millisecondes écoulées depuis le 1er janvier 1970.

7. <http://www.dailymail.co.uk/sciencetech/article-2120416/Twitter-predicts-stock-prices-accurately-investment-tactic-say-scientists.html>

8. <http://singularityhub.com/2010/10/21/twitter-predicts-the-stock-market/>

9. abréviation utilisée pour désigner une action sur les marchés financiers

10. <http://www.infochimps.com/datasets/twitter-census-stock-tweets>

1. La plupart des tweets sont envoyés à cause d'un événement observé dans le monde, sans lien avec d'autres tweets.
2. Un tweet peut générer des nouveaux tweets (une réponse à ce tweet, ou un "retweet"<sup>11</sup>). Certains tweets contiennent une information suscitant beaucoup de réactions. Cela génère un grand nombre de tweets, qui eux-même génèrent un grand nombre de tweet, etc. Dans ce flux, on observe donc une "auto-excitation", qui peut produire des périodes de temps contenant beaucoup de tweets.

L'objectif est d'évaluer la probabilité d'avoir un nombre très élevé de tweets dans une période de temps donnée, en fonction du régime d'auto-excitation considéré, et de la durée de cette période. On cherchera notamment à intégrer dans la modélisation le fait qu'un tweet a un impact sur le flux qui diffère selon son contenu. Les points à traiter sont les suivants.



Tweets (en rose) et retweets (en vert) dans l'heure qui a suivi le tremblement de terre au Japon (mars 2011).  
SOURCE : <http://www.flickr.com/photos/twitteroffice/5884626815/>

1. Simuler le flux de tweet sur une journée, en prenant la seconde comme unité de temps. On paramétrisera le modèle pour avoir, en dehors du phénomène d'auto-excitation, 1 tweet par seconde. Proposer une paramétrisation du modèle permettant de modéliser une auto-excitation faible ou forte, qui persiste longtemps ou diminue rapidement dans le temps.
2. La forte croissance de Twitter a conduit il y a quelques années à la "twitpocalypse"<sup>12 13</sup>, qui est survenue quand le nombre d'identifiants possibles des tweets (qui étaient alors codés avec un entier sur 32 bits) a été dépassé. Evaluer la probabilité d'une twitpocalypse en fonction de la période de temps (jour, semaine, mois), et du niveau d'auto-excitation du modèle.
3. Chaque tweet a des caractéristiques ("features") qui ont un impact plus ou moins fort sur l'auto-excitation du flux. Des exemples de features sont la présence d'un mot, ou d'une combinaison de mots dans le tweet, la présence d'un URL, etc. Proposer une modélisation simple de ce phénomène à l'aide d'une dizaine de features possibles.

Simuler des configurations avec des features qui ont une très faible probabilité d'apparaître dans un tweet (de l'ordre de 0.01%, mais qui ont

11. un utilisateur fait un copier/coller du tweet et le renvoie

12. <http://www.twitpocalypse.com>

13. <http://techcrunch.com/2009/06/12/all-hell-may-break-loose-on-twitter-in-2-hours/>

un impact très fort sur le flux). Simuler également des configurations avec des *features* récurrentes (probabilité de l'ordre de 1%), mais qui ont un faible impact sur le flux.

Déterminer le nombre moyen de tweets apparaissant dans l'heure (ou plus) suivant un tweet contenant une feature avec un fort impact. Evaluer la probabilité d'une twitpocalypse sur différentes périodes de temps, conditionnellement au fait qu'un message contienne une feature avec un fort impact.

## • Une modélisation de l'auto-excitation (sans impact de la feature)

On se place dans un intervalle de temps  $[0, T]$ , l'unité de temps est la seconde, et  $T$  correspond à une journée. On note  $t_1 < t_2 < \dots < t_{N_T}$  la suite des temps d'envoi d'un tweet, tombant dans l'intervalle  $[0, T]$ , où on a posé  $N_t = \sum_{i \geq 1} \mathbf{1}_{\{t_i \leq t\}}$ , le processus de comptage associé à cette suite. On modélise la suite  $\{t_i\}_{i \geq 1}$  par un processus ponctuel "auto-excité", appelé aussi processus de Hawkes [Haw71], en s'inspirant de [SJ12]. Ce type de processus est également très utilisé pour modéliser les tremblements de terre [Oga98]. Dans ce type de modèle, l'intensité du processus de comptage  $\{N_t : t \geq 0\}$  est de la forme

$$\lambda(t) = \lambda_0 + \sum_{i \geq 1, t_i \leq t} g(t - t_i), \quad (1)$$

où  $\lambda_0 > 0$  et  $g$  est une fonction déterministe  $\mathbb{R} \rightarrow \mathbb{R}^+$  telle que  $g(t) = 0$  pour tout  $t \leq 0$ , et décroissante sur  $\mathbb{R}^+$ . L'intensité  $\lambda$  est donc aléatoire, et  $\lambda(t)$  dépend, à l'instant  $t$ , des temps  $t_i < t$ . Cela garantit la propriété de "causalité" : seuls les tweets passés ont une influence éventuelle sur les tweets à venir. La décroissance de  $g$  modélise le fait que l'impact d'un tweet décroît avec le temps. La forme de cette fonction quantifie la "vitesse" de cette décroissance.

L'intensité (1) est la somme des fonctions d'intensité  $\lambda_0, g(\cdot - t_1), g(\cdot - t_2), \dots$ . On peut donc simuler un processus ponctuel d'intensité (1) en "superposant" des générations de processus de Poisson admettant ces intensités. Le processus de comptage d'une génération  $k \in \mathbb{N}$  de temps  $\{t_i^{(k)} : i \geq 1\}$  est noté

$$N_t^{(k)} = \sum_{i \geq 1} \mathbf{1}_{\{t_i^{(k)} \leq t\}}.$$

Le principe de superposition est le suivant.

1. Pour la première génération  $k = 0$ , on simule des temps

$$\Pi_{(0)} = \{t_1^{(0)}, \dots, t_{N_T^{(0)}}^{(0)}\} \sim \text{PP}(\lambda_0),$$

(processus de Poisson d'intensité  $\lambda_0$ ).

2. Puis, on procède par récurrence sur  $k \geq 1$ . On génère les temps de la génération  $k$  à partir de ceux de la génération  $k - 1$  :

$$\Pi_{(k)} = \{t_1^{(k)}, \dots, t_{N_T^{(k)}}^{(k)}\} \sim \text{PP}\left(\sum_{i \geq 1} g(\cdot - t_i^{(k-1)})\right),$$

(processus de Poisson d'intensité  $\sum_{i \geq 1} g(\cdot - t_i^{(k-1)})$ ).

3. On s'arrête à la génération  $K$  telle que  $N_T^{(K)} = 0$  (à chaque étape, on ne rajoute des temps qu'après les temps précédents, donc au bout d'un certain nombre d'itérations, la nouvelle génération sort complètement de l'intervalle  $[0, T]$ ).
4. On construit alors le processus superposé :

$$\{t_1, \dots, t_{N_T}\} = \bigcup_{k=0}^{K-1} \Pi_{(k)},$$

$$\text{où } N_t = \sum_{k=0}^{K-1} N_t^{(k)}.$$

## ● Extension

Estimer les paramètres du processus Hawkes en calculant l'estimateur au maximum de vraisemblance dans le modèle considéré.

## ● Références bibliographiques

Modèle d'auto-excitation et ses propriétés : [\[Haw71\]](#) [\[Oga98\]](#) [\[Bar63\]](#)

Applications et modélisation : [\[Oga98\]](#) [\[SJ12\]](#)

Simulation : [\[Oga81\]](#)

## 7 Microstructure des prix financiers

Les propriétés des trajectoires du prix d'un actif coté sur un marché boursier dépendent fortement de la fréquence d'observation de la trajectoire. Si pour des observations à basse fréquences (échelle de temps : jour, mois) les trajectoires sont bien approchées par des processus de diffusion (par exemple, un mouvement Brownien : mouvement continu d'une particule soumise à des chocs aléatoires gaussiens. Figure 3 (a)), pour des fréquences plus élevées (échelle de temps : heure, minute, ou en dessous), le prix a une nature discrète et est représenté par des trajectoires discontinues, constantes par morceaux qui varient sur une grille, Figure 3 (b). L'amplitude minimale de variation du prix à chaque sauts (0.01 dans la Figure 3) est appelée *tick*.

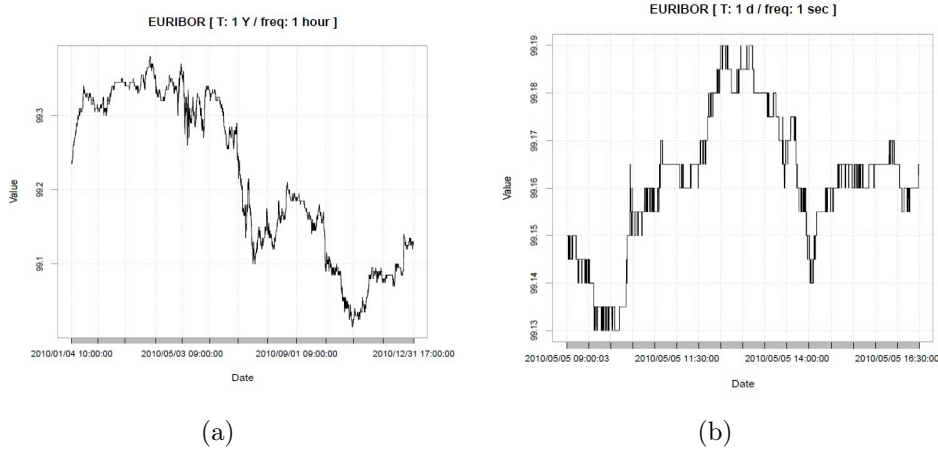


FIGURE 3 – Trajectoires du prix de l'Euribor future pour différentes fréquences d'observation. Pham (2013), présentation au NUS-UTokyo Workshop on Quantitative Finance.

Suivant une approche typique dans la modélisation, on décrit ce prix qui varie sur grille discrète par le biais d'un processus ponctuel  $(T_n, J_n)_{n \geq 1}$ . La suite croissante  $(T_n)_n$  représente les instants de variations (sauts d'un ou plusieurs ticks) du prix. Les v.a.  $J_n$  prennent leurs valeurs dans l'ensemble fini

$$E = \{+1, -1, \dots, +m, -m\} \subset \mathbb{Z} \setminus \{0\}.$$

et représentent les incréments (positifs ou négatifs) du prix. Le prix de l'actif est donné par le processus (constant par morceaux)

$$P_t = P_0 + \sum_{k=1}^{N_t} J_k \quad (2)$$



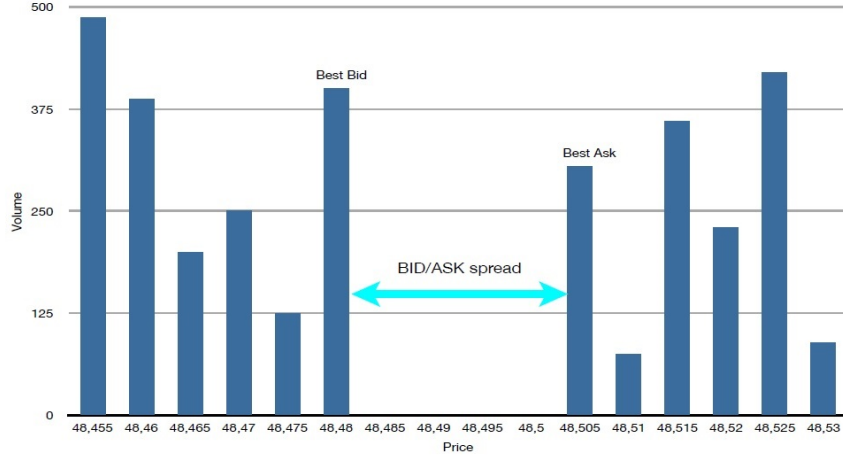


FIGURE 4 – Instantané d'un carnet d'ordre limite (*limit order book*).

où  $N_t$  est le nombre d'occurrences avant  $t$ ,

$$N_t = \sum_{n \geq 1} 1_{T_n \leq t}$$

La valeur courante  $P_t$  représente la dernière cotation du prix *mid*, soit la moyenne entre le meilleur ask et le meilleur bid, Figure 4.

▷ **Une modélisation simplifiée.** Dans le cadre d'une première modélisation simplifiée, on suppose que la suite des incréments  $(J_n)_n$  est i.i.d. (absence de dépendance entre les variations du prix). On suppose de plus que les instants de sauts  $(T_n)_n$  sont indépendants des  $(J_n)_n$ , et que la suite des durées des intervalles entre sauts  $S_n = T_{n+1} - T_n$  est i.i.d. de loi exponentielle de paramètre  $\frac{1}{\lambda}$ .

Q1. On souhaite estimer la probabilité que dans le modèle ainsi défini les trajectoires du prix puissent prendre des valeurs négatives,

$$\mathbb{P} \left( \inf_{t \leq T} P_t < 0 \right)$$

pour  $P_0 = 10$  et  $T = 4h$ . On prendra une valeur de  $\lambda$  telle que la durée moyenne entre un saut et le suivant soit autour de 300s. On considérera les cas  $m = 1$  et  $m = 3$  avec loi des sauts symétrique :  $\mathbb{P}(J_n = 1) = \mathbb{P}(J_n = -1) = 1/2$  pour  $m = 1$  et  $\mathbb{P}(J_n = i) = p_{|i|}$  pour  $i \in E$  et  $m = 3$ , où  $p_1 = 1/2$ ,  $p_2 = 1/3$ ,  $p_3 = 1/6$ .

Pour ces mêmes valeurs des paramètres, on veut estimer les quantiles à  $10^{-4}$ , à gauche et à droite, pour la distribution de  $P_T$ .

▷ **Prise en compte d'autres propriétés des données haute fréquence.**

Il est important qu'un modèle de prix d'actifs à haute fréquence soit capable de reproduire, en plus de la nature discrète des prix, les faits stylisés suivants :

- *Retour vers la moyenne* : les variations du prix sur des courtes échelles de temps sont fortement anti-corrélées. Cela détermine un effet appelé retour vers la moyenne : le prix, tout en étant soumis à des variations aléatoires, a tendance à osciller autour d'une valeur intermédiaire. Ce phénomène est en lien avec la structure du carnet d'ordre : l'exécution d'un ordre à l'achat ou à la vente consomme l'un de deux cotés du carnet. Si tout le volume à disposition au meilleur prix bid ou au meilleur prix ask est entièrement épuisé par l'exécution de l'ordre, l'effet est d'augmenter le best ask, ou de diminuer le best bid (et par conséquent le prix mid change aussi).
- *Clusters de volatilité* : on observe une alternance de périodes avec activité de trading intense (volatilité du prix plus élevée) et de périodes plus calmes (volatilité plus faible). Par conséquent, les instants de sauts ne sont pas distribués de manière uniforme.

► Afin de rajouter une plus forte composante de retour vers la moyenne, on peut modéliser le prix mid par une superposition de processus

$$P_t = P_t^{(1)} + P_t^{(2)},$$

où  $P_t^{(i)} = P_0/2 + \sum_{n=1}^{N_t^{(i)}} J_n^{(i)}$  pour  $i = 1, 2$ , toutes les variables aléatoires indexées par  $i = 1$  et  $i = 2$  étant indépendantes.

(1) : on fera les mêmes hypothèses que dans la section précédente pour les distributions de  $(N_t^{(1)})_{t \geq 0}$  et  $(J_n^{(1)})_{n \geq 1}$ , en changeant la durée moyenne  $1/\lambda_1$  entre les sauts de 300s à 660s.

(2) : la suite de sauts  $J_{n+1}^{(2)} = -J_n^{(2)}$ , à partir de  $J_1^{(2)}$  avec  $\mathbb{P}(J_1^{(2)} = 1) = \mathbb{P}(J_1^{(2)} = -1) = 1/2$ , modélisera la composante soumise à oscillation. On prendra pour  $N_t^{(2)}$  un processus de Poisson simple, de vie moyenne entre sauts  $1/\lambda_2 = 110s$ .

Q2. On veut reprendre les points de la question précédente afin d'estimer l'impacte d'une telle modélisation.

► Une autre approche à la modélisation du phénomène de retour vers la moyenne consiste à s'affranchir de l'hypothèse d'indépendance entre les sauts. On décompose à ce fin la variation du prix à chaque saut en

$$J_n = \hat{J}_n \xi_n$$

où  $\hat{J}_n = \text{sign}(J_n)$  et  $\xi_n = |J_n|$ . On supposera maintenant que

- Les variations du prix ne sont pas indépendantes entre elles, mais ce n'est que le signe d'un incrément de prix (et pas sa valeur absolue) qui influence les incréments futures ; de plus, le signe de l'incrément à l'instant  $T_n$  ne dépend que de celui en  $T_{n-1}$ , et pas des précédents. Mathématiquement, cela se traduit en prenant pour  $\hat{J}_n$  une chaîne de Markov, de matrice de transition

$$\hat{Q} = \begin{pmatrix} \frac{1+\alpha_+}{2} & \frac{1-\alpha_+}{2} \\ \frac{1-\alpha_-}{2} & \frac{1+\alpha_-}{2} \end{pmatrix}$$

où  $\alpha_+, \alpha_- \in [-1, 1)$ .

- Les  $(\xi_n)_n$  sont i.i.d. à valeurs dans  $\{1, \dots, m\}$ , indépendants des  $(\hat{J}_n)_n$ , de loi  $\mathbb{P}(\xi_n = i) = p_i$ ,  $i = 1, \dots, m$ . On modélise de cette manière des amplitudes de saut du prix qui sont indépendantes entre elles, et indépendantes de la direction du saut (hausse ou baisse). Cette propriété traduit une indépendance des volumes échangés dans le carnet d'ordre à chaque événement de trading, et leur indépendance par rapport au type de transaction (vente ou achat).

Q. Quel signe faudra-t-il choisir pour les paramètres  $\alpha_{\pm}$  afin de modéliser le phénomène de retour vers la moyenne ?

On gardera par ailleurs la modélisation (2) pour le prix, avec l'hypothèse d'indépendance entre  $(J_n)_n$  et  $(T_n)_n$ , et une distribution exponentielle pour les  $(S_n)_n$ .

Q3. Des estimations à partir des données sur l'Euribor future [FP15] ont montré qu'il est réaliste de considérer le cas symétrique  $\alpha_+ = \alpha_- = \alpha$ . On fixera pour  $\alpha$  la valeur estimée dans l'article [FP15], Section 2. Simuler des trajectoires du prix  $P_t$ , observer empiriquement sur les simulations que l'on a  $\mathbb{P}(\hat{J}_n \hat{J}_{n+1} = 1) \approx 0.1$ . Estimer dans ce cas également la probabilité que le prix  $P_t$  prenne des valeurs négatives au cours de  $N = 100$  variations et les quantiles à  $10^{-4}$  de la valeur finale du prix.

▷ **Limite macroscopique.** On souhaite étudier le comportement des prix engendré par le modèle considéré dans la question précédente sur des échelles de temps plus longues.

En supposant une durée de  $8h$  pour chaque jour et 250 jours de trading par an, horizon  $T = 1$  an, estimer la probabilité que le prix  $P_t$  prenne des valeurs négatives, ainsi que les quantiles à  $10^{-6}$  de la valeur finale du prix, à partir de la valeur  $P_0 = 1000$  (typique pour un indice boursier) et à un horizon  $T = 1$  an.

## • Références bibliographiques

[FP15]

## 8 Vie et mort d'une population, processus de Yule

**1) Première partie.** Soit  $\lambda > 0$  un paramètre. Partant d'un individu, on s'intéresse à l'évolution d'une population où chaque individu vit un temps aléatoire distribué suivant une loi exponentielle de paramètre  $\lambda$  (indépendamment des autres individus), et meurt en donnant naissance à deux individus. On dit que les individus se reproduisent à un taux  $\lambda$  (on rappelle qu'une loi exponentielle de paramètre  $\lambda$  a pour densité  $\lambda e^{-\lambda x}$  sur  $\mathbb{R}_+$ ). On note  $Y_t$  le nombre d'individus présents à l'instant  $t$  (ainsi,  $Y_0 = 1$  et  $Y$  augmente de 1 à chaque fois qu'un individu meurt). Le processus  $(Y_t)_{t \geq 0}$  est appelé processus de Yule de paramètre  $\lambda$  (il a été introduit par Yule en 1924, mais on l'appelle parfois processus de Yule-Furry car il a été utilisé par Furry pour modéliser une réaction radioactive). Par ailleurs, l'arbre généalogique formé par cette population apparaît en informatique sous le nom de *arbre binaire de recherche* (Binary Search Tree) et y joue un rôle important en analyse des algorithmes. Finalement, pour un entier  $C \geq 1$ , on notera  $Y_t(C)$  le nombre d'individus présents à l'instant  $t$  lorsqu'on part de  $C$  individus à  $t = 0$  (ainsi  $Y_t = Y_t(1)$ ).

▷ **Boîte à outils.** Voici quelques résultats théoriques connus sur les processus de Yule (voir [Nor98, Section 2.5] and [AN72, Section III.7] pour (A), (B) et (D)).

- (A) En pratique, il peut être utile de construire le processus de Yule partant d'un individu comme suit. Considérons une suite de variables aléatoires indépendantes  $(E_i : i \geq 1)$  telle que  $E_i$  suit une loi exponentielle de paramètre  $\lambda_i$  pour tout  $i \geq 1$ . Alors le processus  $(X_t)_{t \geq 0}$  défini par

$$X_t = 1 + \inf\{k \geq 0 : E_1 + E_2 + \dots + E_k \geq t\}.$$

est un processus de Yule de paramètre  $\lambda$  (c'est-à-dire que  $(X_t)_{t \geq 0}$  et  $(Y_t)_{t \geq 0}$  ont la même loi). Dans cette représentation,  $E_i$  est le temps du  $i$ -ième décès. Ceci provient de la propriété d'absence de mémoire de la loi exponentielle, et du fait que le minimum de variables aléatoires exponentielles indépendantes de paramètres respectifs  $\lambda_1, \lambda_2, \dots, \lambda_n$  est une variable aléatoire exponentielle dont le paramètre est  $\lambda_1 + \lambda_2 + \dots + \lambda_n$ .

- (B) Pour tout  $t > 0$  fixé,  $Y_t$  suit une loi géométrique de paramètre  $e^{-\lambda t}$ , c'est-à-dire que  $\mathbb{P}(Y_t = k) = e^{-\lambda t} (1 - e^{-\lambda t})^{k-1}$  pour  $k \geq 1$ .
- (C) Soit  $C \geq 1$  un entier. On pose  $F(n) = \sum_{i=1}^n \ln\left(\frac{i+C}{i}\right)$  et  $L_t = e^{F(Y_t-1) - \lambda C t}$ . Alors  $\mathbb{E}[L_t] = 1$  pour tout  $t \geq 0$ . Ceci permet de définir une nouvelle mesure de probabilité  $\mathbb{Q}$ , et sous  $\mathbb{Q}$  ce nouveau processus a la même loi

que  $(Y_t(C+1) - C)_{t \geq 0}$  ([de 06]). Autrement dit, pour toute fonction positive  $g$ ,

$$\mathbb{E}_{\mathbb{P}}[g(Y_t)] = \mathbb{E}_{\mathbb{Q}} \left[ g(Y_t) \cdot \frac{1}{L_t} \right] = \mathbb{E}_{\mathbb{P}} \left[ g(Y_t(C+1) - C) \cdot \frac{1}{e^{F(Y_t(C+1) - C - 1) - \lambda C t}} \right].$$

- (D) Lorsque  $t \rightarrow \infty$ ,  $Y_t \cdot e^{-\lambda t}$  converge presque sûrement vers une variable aléatoire notée  $\mathcal{E}$  et appelée *valeur terminale* du processus de Yule  $(Y_t)_{t \geq 0}$ . Il se trouve que  $\mathcal{E}$  suit une loi exponentielle de paramètre 1.
- (E) Il peut être utile de simuler un processus de Yule ayant une valeur terminale fixée. Plus précisément soit  $\mathcal{E}$  une variable aléatoire exponentielle de paramètre 1 et soit  $(P_t)_{t \geq 0}$  un processus de Poisson indépendant homogène de paramètre 1. Alors le processus  $(Z_t)_{t \geq 0}$  défini par

$$Z_t = 1 + P_{\mathcal{E}(e^{\lambda t} - 1)}$$

est un processus de Yule de paramètre  $\lambda$  et de valeur terminale  $\mathcal{E}$  ([AN72, Section 11 dans Chapter III, Theorem 2] ou [Ken66, Theorem 1]).

#### ▷ Questions.

- (1) Simuler les valeurs de  $\mathbb{P}(Y_t > A)$  et de  $\mathbb{P}(Y_t < A)$ , en particulier dans le régime des événements rares. Comparer les différentes méthodes utilisées.
- (2) Illustrer la convergence du point (D) par des simulations. Étudier par des simulations l'influence de la valeur terminale  $\mathcal{E}$  sur la loi de  $Y_t$ , en particulier dans le régime des événements rares.

**2) Deuxième partie.** Dans cette partie, on suppose que la population peut se déplacer spatialement et qu'elle est chassée par des prédateurs. Voici un modèle simplifié :

- Chaque site de  $\mathbb{N} = \{0, 1, 2, \dots\}$  est occupé soit par une seule proie, un seul prédateur ou est vide.
- A l'instant  $t = 0$ , un prédateur occupe le site 0, une proie occupe le site 1 et tous les autres sites sont vides.
- Si une proie est à côté d'un site vide, au bout d'un temps aléatoire, indépendant de tout le reste, distribué selon une variable exponentielle de paramètre  $\lambda$ , elle donne naissance (sans mourir) à une nouvelle proie qui vient occuper ce site vide.
- Si un prédateur est à côté d'une proie, au bout d'un temps aléatoire, indépendant de tout le reste, distribué selon une variable exponentielle de paramètre 1, il donne naissance à un nouveau prédateur qui vient occuper le site occupé par la proie.

On note  $p(\lambda)$  la probabilité que les proies disparaissent. Il est possible de démontrer que  $p(\lambda) = 0$  pour  $\lambda \leq 1$  et que  $p(\lambda) > 0$  pour  $\lambda > 1$ .

Un outil efficace pour étudier ce modèle est le processus  $(S_n : n \geq 0)$ , où  $S_n$  est le nombre de proies après  $n$  naissances (soit de proies, soit de prédateurs). Ainsi,  $S_0 = 1$  et les proies disparaissent s'il existe un entier (aléatoire)  $k$  tel que  $S_k = 0$ . On pourra vérifier que  $(S_n : n \geq 0)$  est une marche aléatoire dont on calculera la loi de saut.

▷ **Questions.**

- (3) Tracer numériquement le graphe de  $\lambda \mapsto p(\lambda)$  pour  $\lambda > 1$ , et étudier numériquement si la fonction  $p$  est dérivable à droite en 1. Lorsque  $\lambda > 1$ , on pourra admettre qu'on peut approximer l'événement « les proies ne disparaissent pas » par l'événement « les proies atteignent le site  $k$  » quitte à faire une erreur de  $\frac{\lambda-1}{\lambda^{k+1}-1}$  sur  $p(\lambda)$ .
- (4) Notons  $A_n(\lambda)$  la probabilité qu'une proie atteigne le site  $n$ . Lorsque  $\lambda < 1$ , estimer  $A_n(\lambda)$  pour quelques valeurs de  $n$  et de  $\lambda$  dans le régime des événements rares. On pourra en particulier prendre  $(\lambda, n) = (0.95, 100)$ ,  $(\lambda, n) = (0.6, 40)$ ,  $(\lambda, n) = (0.5, 10)$  et  $(\lambda, n) = (0.2, 7)$ .
- (5) Lorsque  $\lambda < 1$ , on peut démontrer que  $A_n(\lambda) \sim a(\lambda) \cdot b(\lambda)^n$  lorsque  $n \rightarrow \infty$  pour certaines valeurs  $a(\lambda), b(\lambda) > 0$ . Tracer numériquement les graphes des fonctions  $a$  et  $b$ . Que se passe-t-il pour  $\lambda = 1$  ?

**3) Troisième partie.**

▷ **Question.**

- (4) Reprendre la deuxième partie en travaillant maintenant sur  $\mathbb{Z}^2$ , avec à l'instant  $t = 0$  une proie et un prédateur côte à côte, et tous les autres sites vides.

● **Références bibliographiques**

[AN72][Ken66][Nor98][de 06]

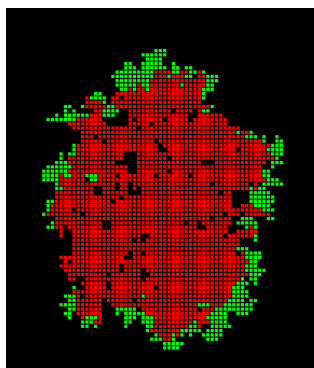


FIGURE 5 – Simulation de la population sur  $\mathbb{Z}^2$ , où les sites verts correspondent au proies et les sites rouges aux prédateurs.

## Références

- [AN72] K.B. Athreya and P.E. Ney. *Branching processes*. Springer-Verlag, New York, 1972.
- [Asm03] S. Asmussen. *Applied probability and queues*, volume 51 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.
- [Bar63] M.S. Bartlett. The spectral analysis of point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 264–296, 1963.
- [BB09] V. Bansaye and J. Berestycki. Large deviations for branching processes in random environment. *Markov Process. Related Fields*, 15(4) :493–524, 2009.
- [BLG09] J.H. Blanchet, K. Leder, and P.W. Glynn. Efficient simulation of light-tailed sums : an old-folk song sung to a faster new tune. . . . In *Monte-Carlo and quasi-Monte-Carlo methods 2008*, pages 227–248. Springer, Berlin, 2009.
- [BR78] J. Bussac and P. Reuss. *Traité de neutronique*. Hermann, Paris, 1978.
- [CC10] R. Carmona and S. Crépey. Particle methods for the estimation of credit portfolio loss distributions. *Int. J. Theor. Appl. Finance*, 13(4) :577–602, 2010.
- [CDL<sup>+</sup>89] M. Cessenat, R. Dautray, G. Ledanois, P.L. Lions, E. Pardoux, and R. Sentis. *Méthodes probabilistes pour les équations de la physique*. Collection CEA, Eyrolles, 1989.
- [de 06] A. de La Fortelle. Yule process sample path asymptotics. *Electron. Comm. Probab*, 11 :193–199, 2006.
- [DG05] P. Del Moral and J. Garnier. Genealogical particle analysis of rare events. *Annals of Applied Probability*, 15 :2496–2534, 2005.
- [FP15] P. Fodra and H. Pham. Semi-markov model for market micro-structure. *Applied Mathematical Finance, Preprint arXiv <http://arxiv.org/abs/1305.0105>*, 22(3) :261–295, 2015.
- [FS02] H. Föllmer and A. Schied. *Stochastic finance*, volume 27 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2002. An introduction in discrete time.
- [GHS00] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Variance reduction techniques for estimating value-at-risk. *Management Science*, 46 :1349–1364, 2000.



- [GHS02] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Portfolio value-at-risk with heavy-tailed risk factors. *Math. Finance*, 12(3) :239–269, 2002.
- [GKV03] J. Geiger, G. Kersting, and V. Vatutin. Limit theorems for subcritical branching processes in random environment. *Ann. Inst. H. Poincaré Probab. Statist.*, 39(4) :593–620, 2003.
- [GL04] W. Glover and J. Lygeros. A multi-aircraft model for conflict detection and resolution algorithm evaluation. *HYBRIDGE WP1 : Identification and modelling of uncertain hybrid systems*, 2004.
- [Gla03] P. Glasserman. *Monte-Carlo methods in Financial Engineering*. Springer Verlag, New York, 2003.
- [Haw71] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1) :83–95, 1971.
- [HPS05] J. Hu, M. Prandini, and S. Sastry. Aircraft conflict prediction in the presence of a spatially correlated wind field. *IEEE Transactions on Intelligent Transportation Systems*, 6(3) :326–340, 2005.
- [JY98] V.A.A. Jansen and J. Yoshimura. Populations can persist in an environment consisting of sink habitats only. *Proc. Natl. Acad. Sci. USA, Ecology*, 95 :3696–3698, 1998.
- [KAT<sup>+</sup>08] D.A. Keith, H.R. Akçakaya, W. Thuiller, G.F. Midgley, R.G. Pearson, S.J. Phillips, H.M. Regan, M.B. Araújo, and T.G. Rebelo. Predicting extinction risks under climate change : coupling stochastic population models with dynamic bioclimatic habitat models. *Biol. Lett.*, 4 :560–563, 2008.
- [KBCP99] C. Kollman, K. Baggerly, D. Cox, and R. Picard. Adaptive importance sampling on discrete Markov chains. *Annals of Applied Probability*, 9(2) :391–412, 1999.
- [Ken66] D.G. Kendall. Branching processes since 1873. *Journal of the London Mathematical Society*, 1(1) :385–406, 1966.
- [Lag06] A. Lagnoux. Rare event simulation. *Probab. Engrg. Inform. Sci.*, 20(1) :45–66, 2006.
- [Lag08] A. Lagnoux-Renaudie. Effective branching splitting method under cost constraint. *Stochastic Process. Appl.*, 118(10) :1820–1851, 2008.
- [LES03] R. Lande, S. Engen, and B.E. Saether. *Stochastic population dynamics in ecology and conservation*. Oxford University Press, Oxford, 2003.
- [LL91] D. Lambertson and B. Lapeyre. *Introduction au calcul stochastique en Finance*. Collection Mathématiques et Applications 9 - Ellipses, 1991.

- [LPS98] B. Lapeyre, E. Pardoux, and R. Sentis. *Methodes de Monte-Carlo pour les processus de transport et de diffusion*. Collection Mathématiques et Applications 29 - Springer Verlag, 1998.
- [M3A] S. Méléard. *Modèles aléatoires en Ecologie et Evolution*. Cours de l'Ecole Polytechnique, MAP563, 3A.
- [Nel67] E. Nelson. *Dynamical theories of Brownian motion*. Princeton University Press, Princeton, N.J., 1967.
- [Ner77] O. Nerman. On the maximal generation size of a non-critical Galton-Watson process. *Scandinavian Journal of Statistics*, 4(3) :131–135, 1977.
- [Nor98] J.R. Norris. *Markov chains*, volume 2. Cambridge university press, 1998.
- [Oga81] Y. Ogata. On lewis' simulation method for point processes. *Information Theory, IEEE Transactions on*, 27(1) :23–31, 1981.
- [Oga98] Y. Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2) :379–402, 1998.
- [Pap75] G. Papanicolaou. Asymptotic analysis of transport processes. *Bulletin of the American Mathematical Society*, 81(2) :330–392, 1975.
- [PE97] R.A. Paielli and H. Erzberger. Conflict probability estimation for free flight. *Journal of Guidance, Control, and Dynamics*, 20(3) :588–596, 1997.
- [PW05] M. Prandini and O.J. Watkins. Probabilistic aircraft conflict detection. *HYBRIDGE WP3 : Reachability analysis for probabilistic hybrid systems*, 2005.
- [SJ12] A. Simma and M.I. Jordan. Modeling events with cascades of Poisson processes. *arXiv preprint arXiv :1203.3516*, 2012.
- [TT3A] P. Tankov and N. Touzi. *Calcul stochastique en finance*. Cours de l'Ecole Polytechnique, MAP552, 3A.